

REPORTER GUIDES

Making Sense of Education Research

BY DEBRA VIADERO

Most education reporters from time to time will tread into the world of education research, whether to gauge charter school achievement, the impact of teacher quality, or the effects of a reading program, among myriad possibilities. But making sense of the research, with its often-impenetrable prose, dizzying figures, and mathematical formulas, can be daunting. Despite the challenge, gaining some basic skills and knowledge in navigating research makes for stronger journalism.

Research in education serves a wide variety of functions. It can be used to assess the likelihood that a new policy or practice will be effective, raise questions about possible unintended side effects of such changes, or shed light on students' and teachers' behaviors. Research can also muddle debates – and sometimes unfairly cast aspersions – on well-intended strategies for improvement. The trick for education reporters is to understand when research findings are trustworthy and when they are not. Reporters need to know how to assess research findings, where to find credible research, and what the limitations and red flags are in the studies they encounter.

GETTING HELP

Research comes across education reporters' radar screens in at least two ways. The most obvious is when a major study is issued that speaks to an important question of education policy or practice. The second is when, in the course of covering news on the beat, the reporter wonders: "What does the research say on that?"

Either way, the reporter's job is to know where to find the relevant studies, how to understand them, how to assess their significance and credibility, and how to clearly communicate that information to the public.

For local reporters, a good place to get help is the education school at a nearby college or university. The scholars there can help translate findings and identify colleagues across the country who are doing similar work.

At the national level, organizations such as the American Educational Research Association, the American Sociological Association, and the American Psychological Association can direct reporters to academics who are studying the same topic and may be willing to comment on the research. Indeed, a large and growing number of education researchers have come to understand the value of speaking with reporters, and the communications teams at some major universities have become sophisticated in their outreach when timely studies are issued.

Within a particular study, the footnotes and appendix provide a

useful road map to the names of other scholars who can lend a hand in vetting the findings. They will be listed as primary authors on the studies cited in the research article.

With help from a good librarian, reporters may also be able to find useful, relatively recent meta-analyses to provide background on their research questions. For a meta-analysis – essentially, a study of studies – researchers may look at dozens and sometimes hundreds of academic reports on the same question in order to calculate the overall effect of a particular intervention or program. Meta-analyses in education, for example, have addressed the academic impacts on students of homework, school uniform policies, and peer tutoring, among other topics, and some have become touchstones for anyone looking to get the lay of the land for those issues. The authors of those meta-analyses are arguably the most authoritative sources to speak on "what the research says" in their study area.

KEY QUESTIONS

It can be comforting to think of research as the ultimate authority on a question of educational policy or practice, but the truth is that usually it is not. The best that research can do is to provide clues on what works, when, and for whom, because



classrooms, schools, and communities inevitably vary. A reading program that achieved good results in an affluent Fairfax County, Va., community might not work so well in a rural, Mississippi Delta school, for instance.

At worst, research can provide a skewed take on what it is examining, as sometimes happens when advocacy groups attempt to spin a public debate by producing their own head-line-grabbing studies.

Reporters should ask some key questions to help understand the limitations of studies and sift out credible nuggets of research information from findings that are suspect. Also, keep in mind that it is always best to go straight to the source and read the full study.

- Who paid for the study? Journalism 101 teaches reporters to be suspicious of information generated by anyone with a stake in the results. This policy holds true for education research, too, generally speaking. But it is also important to be aware that it can be difficult for programs to get outside funding to evaluate their initiatives. And not every study from an advocacy group is necessarily biased or of poor quality. Some advocacy-oriented think tanks produce very credible research; some do not. Others produce a mixture of bad and good. A closer look at the study itself (and perhaps finding another expert to take a look) should be the deciding factor on its merit.
- Where was the study published? In terms of trustworthiness, research published in a peer-reviewed journal almost always trumps research that is published without extensive review from other scholars in the same field.
- How were participants selected for the study? Reporters should always be on the lookout for evidence of "creaming" in

Reporters should ask some key questions to help understand the limitations of studies and sift out credible nuggets of research information from findings that are suspect.

other words, choosing the best and brightest students for the intervention group.

- How were the results measured? It is not enough to state that students did better in reading, math, or another subject. Reporters need to know exactly what measuring stick was used. Was it a standardized test or one that was developed by the researchers? Did the test measure what researchers were actually studying? Did students take the same test before and after the study? (In other words, could students have done better the second time around just because they were older, or more familiar with the test?)
- Was there a comparison group? Reporters should be wary of conclusions based on a simple pre- and post-test conducted with a single group of students. Whether the study design is a random assignment, a regression-discontinuity analysis, or a basic comparison, there should always be a control group with demographic and other characteristics that are similar to those of the experimental group.
- How many participants were involved in the study? Finding out the sample size is key to understanding if the study was

EFFECT SIZE

An effect size measures the magnitude of a research finding. At its most basic level, the effect size is the difference between those who received a particular intervention (the "treatment group") and those who did not (the "control group").

For example, let's say a researcher is studying the effects of a new online course that prepares students for the ACT college-admissions exam. Half of the 11th graders in a school district take the course. Their average score is 22. Half the 11th graders in the district do not take the course. Their average score is 21. So the effect size of the course is 22 minus 21, which equals 1 point on the ACT scale. When reporting on studies of interventions like this hypothetical ACT course, you should always try to include the effect size. That helps

audiences evaluate whether the program produced meaningful results that were worth the time, effort and resources required to implement the program.

If researchers report effect sizes in ways that are difficult to understand (e.g. "5 percent of a standard deviation"), ask them to translate the effect size into units (like points on the ACT scale) that you think your audience will recognize. Also, ask them to explain how the effect size of their study compares to the effect sizes identified by other studies of similar programs or programs that might be used in similar ways. To return to the hypothetical ACT study, you might ask the researchers about the average effect size for online ACT courses studied by other researchers.

– Holly Yettick



large enough for researchers to consider the results statistically significant. (See sidebar on this page.)

- What else happened during the study period that might explain the results? For example, were there any changes in the school's enrollment, teaching staff or leadership?
- What are the limitations of the study? Most good researchers will include a section in their research article that outlines the limitations of the study. This is required reading for reporters who cover the study.
- Are the findings in sync with other research in the same area? If conclusions are different, it's not necessarily a red flag, but it does signal the need to dig deeper. What was different about this experiment compared with previous studies on the same research question?

Other factors also may be important to keep in mind when considering the value of research. For example, was the experimental group really getting a different treatment? In a school-based study, teachers may not have implemented an intervention faithfully. What do the researchers know about "implementation fidelity" in their study?

Likewise, reporters should learn what sorts of instructional treatments students in the control group were getting. A federal evaluation of the short-lived Reading First program found that it led to no sizeable reading gains, but it also determined that students in the comparison group were not being taught in a vacuum. The instructional approaches used in the experimental group were widespread, and, in many cases, the students in the control group were being taught by some of the same methods even though they were not participating in a Reading First program.

STUDY DESIGN

For a time, the conventional wisdom around education research was that only true experiments - studies in which participants were randomly assigned to either a control or a treatment group – could provide reliable, cause-and-effect conclusions. While they are still considered the "gold standard" for establishing causation, such randomized, controlled studies are not infallible. Those experiments can be very expensive to carry out - and even unethical in some cases. If researchers really believe an intervention will improve learning, for example, withholding it from a control group of students could be seen as unjustified.

Regression-discontinuity designs are a type of analysis sometimes used to evaluate interventions that involve some sort of a test-score cutoff – for example, a grade-retention policy that requires students to repeat a grade if their reading scores on a

STATISTICAL SIGNIFICANCE

An important concept in education research is what is known as a "p value." This is the outcome of a significance test that is based on the pessimistic assumption that a researcher's results are merely due to chance. Big values of p are evidence supporting the idea that the results were due to chance. Small values of *p* are evidence against the idea that chance explains the outcome.

As you might imagine, researchers usually hope to obtain small p values. If p is less than 5 percent, the result is often called "statistically significant" because it means there is only a 5 percent chance of seeing a result at least as extreme as the one the researcher has obtained if the results are actually the result of chance. If p is less than 1 percent, the results may be called "highly statistically significant."

The p value is not only among the most widely used statistical terms; it is also among the most misused. One common misuse involves using *p* values in studies in which no chance process occurred, as when a test is given to every single student in a school district instead of to a random selection of students. The p value cannot tell you whether your results were affected by factors unrelated to chance – such as the fact that some of the students in the district were absent the day of the test. After all, the students probably weren't randomly selected to miss school.

Another common misuse is to assume that statistical significance is equivalent to obtaining big, meaningful or important results. Unfortunately, a p value will not really tell you anything about the magnitude or realworld importance of the results. It is important to keep in mind, for instance, that *p* values are very sensitive to sample sizes. The bigger the sample, the more likely the findings are statistically significant. But even while studies with many thousands of participants may yield highly statistically significant results, the actual effect of the program or intervention may be marginal.

– Holly Yettick

standardized test fall below a specified point. The assumption is that students who fall just above or below the cutoff are not all that different from one another, academically speaking. But more-typical, quasi-experimental studies, in which researchers compare a group of students who are getting an experimental intervention with those who are not, can provide good information as well if they are carefully done. The key

2015 © EDUCATION WRITERS ASSOCIATION



ingredient is having a comparison group that looks a lot like the experimental group in terms of students' academic backgrounds, ages, and demographics.

FINDING RESEARCH

Unlike in the medical field, which has a handful of must-read academic journals such as the Journal of the American Medical Association or Lancet, education research is disseminated through dozens of journals. Many published studies also tend to be fairly old, due to the long lag between the time an article is submitted for consideration and actual publication.

Sometimes, an easier way to stay on top of new and newly published studies is academic conferences and news aggregators like Newswire, which disseminates press releases generated by university-based public relations offices. Reporters can also get on the email lists for private research organizations, such as Mathematica or the American Institutes of Research, that do extensive work in education.

If the reporter's aim is to find other studies on the same topic, the What Works Clearinghouse operated by the U.S. Department of Education is one place to look. Researchers can submit studies to the clearinghouse, where they are vetted by department reviewers at the Institute for Education Sciences,

If the reporter's aim is to find other studies on the same topic, the What Works Clearinghouse operated by the U.S. Department of Education is one place to look.

the department's main research arm. The resulting reviews are posted on the What Works website. Be warned: Due to the rigorous methodology used by the department, few studies get an unqualified thumbs-up.

The research agency also periodically publishes intervention reports, which use the same tough What Works standards to assess the body of research evidence on the effectiveness of particular instructional approaches or commercial instructional programs (as opposed to the quality of a specific study on that intervention).

The institute also publishes best-practice reports, which take a broader approach to assessing empirical evidence. Those reports tend to be most useful for practitioners looking for best bets on what might work in their own classrooms, schools, or districts.

for more
REPORTER
GUIDES
go to EWA.org

About the Authors: Debra Viadero is an assistant managing editor at Education Week who oversees the newspaper's coverage of education research and other issues. Holly Yettick is the director of the Education Week Research Center.

This guide was made possible by a grant from the Spencer Foundation.

